

Gestión de Calidad de Datos en la Combinación de Actividades dentro del Marco de los Procesos de Negocio

Isabel Bermejo¹, Luisa Parody², Ismael Caballero¹, María Teresa Gómez-López², and Rafael M. Gasca²

¹ Instituto de Tecnologías y Sistemas de Información, Universidad de Castilla La-Mancha, Paseo de la Universidad 4, 13071, Ciudad Real (España),
{Isabel.Bermejo, Ismael.Caballero}@uclm.es

² Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Sevilla
Avda. de la Reina Mercedes s/n, 41012, Sevilla (España)
{lparody, maytegomez, gasca}@us.es

Abstract. Los procesos de negocio ofrecen la infraestructura necesaria para la combinación de distintas actividades en un único proceso. Dicha combinación puede implicar el intercambio de datos entre las mismas, de forma que la calidad de datos toma una especial relevancia. Si los datos que utilizan las actividades involucradas no tienen el nivel de calidad adecuado, el resultado generado por el proceso podría no ser ni fiable ni usable por el usuario final. Una forma de garantizar la confiabilidad en los datos es mediante una certificación de su nivel de calidad. La certificación de los niveles de calidad de los datos que deben manejar las actividades que se combinan, puede ser descrita mediante el uso de la familia de estándares ISO/IEC 8000-100. El uso de esta certificación supone una nueva restricción que ha de tenerse en cuenta a la hora de buscar el resultado global en la combinación de actividades. En este artículo se propone el diseño y la implementación de una Arquitectura de Servicios llamada I8K, que se encarga del *proceso de evaluación y certificación de los datos*, y de como ésta se ha aplicado a un ejemplo motivador sobre la combinación de actividades para la organización de un viaje, que implica la búsqueda de billete de avión, estancia en hotel, y opcionalmente el alquiler de coches.

Keywords: Calidad de Datos, ISO 8000-100, Arquitectura de Servicios, Procesos de Negocio, Combinación de Actividades

1 Introducción

Redman afirmó que datos, con niveles adecuados de calidad, añaden un gran valor a las actividades de las que se compone el negocio de las organizaciones [21]. Desafortunadamente, pese a que la información gestionada en estas actividades es de gran importancia, gran parte de las organizaciones no tienen en cuenta la calidad de sus datos, y por tanto, no son capaces de lograr ventajas

estratégicas. Sin embargo, datos que no se utilizan o que son incorrectos, hacen que la mayoría de las operaciones y procesos de toma de decisiones fracasen e impidan a las organizaciones conseguir sus objetivos [20]. Malas decisiones pueden desencadenar problemas técnicos, legales u operativos, representando así una verdadera amenaza para el negocio de las organizaciones [9].

Los procesos de negocios (Business Process, BP) tienen como base la representación explícita de los procesos como un conjunto de actividades y las restricciones de ejecución entre ellas [23]. De igual manera, la gestión de procesos de negocios (Business Process Management, BPM), puede ayudar a diferentes organizaciones a volverse más competitivas mediante la asociación de sus actividades con el fin de ofrecer un objetivo global. La adecuada ejecución de estas actividades depende fundamentalmente de los datos de entrada y salida que manejan. Por otra parte, las actividades de los BP pueden apoyarse en el uso de servicios web, los cuales juegan un rol muy importante dentro de BPM, ya que delegan la implementación de la actividad a un componente externo. Las organizaciones proporcionan los servicios con el objetivo de ofrecer ciertas funcionalidades a los usuarios. El problema es que estos servicios no están preparados para interactuar con otros servicios. Por esta razón, el uso de los servicios web se integra fácilmente con los BPM. Por una parte, los BP, mediante la combinación de actividades, ofrecen el soporte necesario para que los servicios web intercambien datos para conseguir un objetivo global. Y por otra parte, los servicios web ofrecen funcionalidades implementadas de manera externa. El problema es que, si el nivel de calidad de los datos intercambiados es insuficiente, el resultado global generado por el proceso completo, tampoco tendrá el nivel de calidad adecuado.

La calidad de datos toma especial relevancia cuando la combinación de las actividades se hace con el objetivo de encontrar la combinación óptima de los datos, por ejemplo minimizar un precio final, maximizar beneficios, reducir tiempo. Parody et al. en [18] muestran que al combinar actividades, se pueden producir esquemas de ejecución no viables cuando se busca el objetivo global óptimo, ya que hay combinaciones que semánticamente no son posibles. Para conseguir el objetivo global óptimo mediante un esquema de ejecución viable, se requieren algoritmos de optimización que contemplen ciertas restricciones semánticas que relacionan los datos de entrada/salida de las diferentes actividades. Por este motivo, a la dificultad propia de los algoritmos de búsqueda de soluciones óptimas con información distribuida (entre los distintos servicios para este caso), se añaden los problemas de la gestión de la certificación de los niveles de calidad en los datos. Por esta razón, en este artículo se propone la infraestructura necesaria para resolver estos problemas.

Para proporcionar la información de la certificación de los niveles de calidad contemplados en las restricciones, en este artículo se usa la familia de estándares ISO/IEC 8000-100, que proponen usar un formato preestablecido para el intercambio de datos y los requisitos que deben cumplir dichos mensajes para incorporar información sobre el nivel de calidad de los datos. Actualmente, no existen muchas implementaciones utilizables públicamente del estándar ISO/IEC 8000:2011, tan sólo la desarrollada por ECCMA, que está accesible en [10]. Sin

embargo, ECCMA sólo implementa las partes 100 y 110 del estándar, mientras que la arquitectura I8K, que se propone en este trabajo, contempla no sólo esas dos, sino también el resto: 120, 130 y 140.

Como parte de la implementación de ISO/IEC 8000-100 se ha desarrollado una Arquitectura de Servicios, llamada **I8K**, que contiene los agentes necesarios para el proceso de evaluación y certificación de la calidad de los datos en los mensajes de datos.

1.1 Ejemplo ilustrativo

Un ejemplo claro de combinación de actividades basadas en servicios es la organización de viajes con el objetivo de minimizar el precio y en función de las preferencias del usuario [19]. El proceso de negocio completo es el que se muestra en la Fig. 1. En primer lugar se recibe la petición del usuario con las preferencias del viaje. Después, se realiza la búsqueda de la mejor combinación de valores para el avión, hotel y alquiler de coche que formarán el paquete de viaje; dichos valores han sido solicitados a los correspondientes servicios web que implementan las actividades. Una vez realizada la búsqueda, se le ofrece al usuario los paquetes de viajes candidatos (que por la naturaleza del algoritmo que los genera son de por sí soluciones óptimas) para que decida si quiere aceptar algún paquete de viaje. Finalmente, según lo que decida, la reserva se cancela o formaliza.

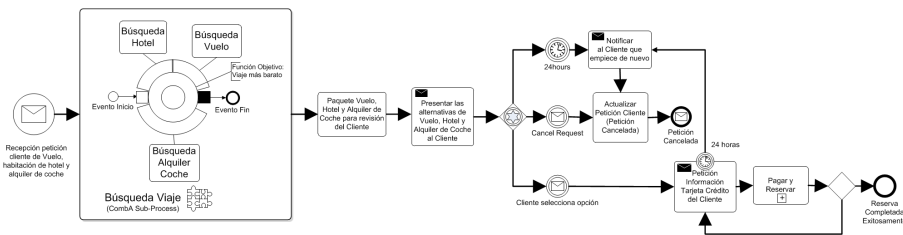


Fig. 1. Ejemplo de Organización de Viaje [18].

Hasta el momento, el usuario proporcionaba únicamente los datos necesarios para organizar un viaje: fechas de salida y regreso, y las ciudades origen y destino. Además, para flexibilizar la búsqueda, el usuario también proporcionaba, unos márgenes de fecha, y un radio de kilómetros (para poder ir a otros aeropuertos) con el fin de encontrar el viaje más barato. Teniendo en cuenta estos datos y las restricciones semánticas existentes para el intercambio de estos entre las distintas actividades, el organizador realiza las llamadas necesarias a los distintos servicios para obtener la información de cada una de las partes (precio del vuelo, precio de la habitación de hotel y precio del alquiler del coche). A continuación, el organizador diseña la posible solución global y la evalúa en función del objetivo global, que en este trabajo es minimizar el precio total del viaje.

Sin embargo, a no ser que los servicios no den la información detallada, hay aspectos que al combinar las actividades no se pueden tener en cuenta. Un ejemplo de este problema se encuentra en la reserva de los billetes de avión: puede

ocurrir que el precio de las maletas no esté incluido en el precio del billete, siendo informado el usuario justo cuando va a formalizar la compra. En este caso, exigir al proveedor de vuelos que informe sobre los servicios que están incluidos en el precio del vuelo (maleta, seguro,...).

Una forma de controlar este tipo de problemas es exigiendo a los servicios que proporcionen unos datos con un nivel de calidad adecuado. De esta manera, el usuario puede saber qué margen de error se puede encontrar en un servicio concreto, por ejemplo, con respecto al precio. Proveer un nivel de calidad de datos adecuado puede ser de tal importancia para los servicios, que pueden llegar a ser incluso excluidos en la búsqueda de la mejor combinación.

1.2 Objetivo del artículo

El trabajo que se presenta se centra en mejorar la eficacia de los algoritmos de búsqueda de combinaciones de actividades. Esta mejora incluye la adición de nuevas restricciones basadas en el nivel certificado de calidad de datos. Además, la mejora propuesta se aplicará al ejemplo presentado en la Sección 1.1 anterior, de modo que sea posible encontrar paquetes de viajes que no sólo cubran las preferencias de los usuarios, sino que además, colmen sus expectativas. Así, cuanto más válido, correcto y usable sea el resultado propuesto, elegido y consumido por los usuarios, mayor será el éxito del proceso, y por tanto de las organizaciones que ofrecen sus servicios en él.

Para presentar estas ideas, el resto del artículo está compuesto de las secciones que se detallan a continuación. En la Sección 2 se describen algunos de los trabajos relevantes sobre la optimización de resultados en los procesos de negocio, así como sobre los aspectos de calidad de datos, incluyendo una descripción de la familia de estándares ISO/IEC 8000-100. En la Sección 3 se presenta la arquitectura de servicios propuesta. La Sección 4 expone las mejoras que se han realizado sobre el algoritmo que es implementado en la aplicación, mostrados en la sección 1.1, donde se incluyen las primitivas de los servicios de la arquitectura. Finalmente, en la Sección 5 se muestran las conclusiones.

2 Estado del Arte

Para mostrar nuestra propuesta es necesario analizar distintas áreas. Por una parte es necesario profundizar en la combinación de actividades basadas en servicios. Respecto a la calidad de datos, se presenta un análisis de los aspectos más importantes en la Gestión de Calidad de Datos y Gestión de Datos Maestros, y por último, se describe brevemente la familia de estándares ISO/IEC 8000-100.

2.1 Combinación de actividades basadas en servicios

Los sistemas orientados a servicios surgieron como paradigma para proporcionar el soporte automático a los BP. Van der Aalst en [22] y Papazoglou [17] presentaron los servicios web como la infraestructura que fomenta la composición de servicios web individuales para representar procesos más complejos.

Los problemas de optimización de restricciones distribuidas describen sistemas donde la información no está centralizada y es conocida por una única entidad, sino que está distribuida entre distintos nodos independientes. Con diferentes protocolos y algoritmos propuestos en una extensa bibliografía, tales como [24],[11] y [8], se obtiene la combinación de los datos que comparten los distintos nodos para optimizar una función dada. Este tipo de problemas se puede extrapolar a los BP, donde cada uno de los nodos representa las actividades basadas en servicios, y los datos compartidos corresponden a los valores que fluyen (data flow) en el proceso.

En los trabajos [19] y [18], Parody et al. presentan ejemplos de cómo combinar actividades basadas en servicios en un BP, con el objetivo de buscar el mejor esquema de ejecución que genera todas las combinaciones de valores de los datos que optimizan el objetivo común. Esta nueva forma de ver la combinación de actividades, se centra en la búsqueda orientada a los datos obtenidos en cada llamada a los servicios. En esta combinación de actividades, el papel de los datos y las restricciones semánticas entre ellos son fundamentales y críticos, ya que los datos de entrada de las diferentes actividades dependen de los datos de salida de las otras actividades, y por tanto son pieza fundamental en la optimización del resultado global obtenido. Esta dependencia, hace que las decisiones tomadas por un servicio influya en las decisiones del resto, y por lo tanto, en el resultado final. En este tipo de problemas es especialmente importante la calidad de los datos intercambiados entre los servicios. Cuanto más fiables y completos sean los datos ofrecidos por los diferentes servicios, mejor será el resultado que se le ofrezca al usuario.

2.2 Gestión de Calidad de Datos Maestros

La definición más aceptada de **calidad de los datos** es *fitness for use* [7]. Es decir, los datos como materias primas deben tener características para que, puedan ser utilizados en un determinado proceso de negocio.

Las organizaciones, a lo largo de su vida, han ido recopilando una gran cantidad de datos que han tenido que adaptar previamente para que les sean útiles y poder así interpretarlos e intercambiarlos con otras aplicaciones y/u organizaciones [14]. Estos datos describen objetos clave de sus negocios y reciben el nombre de **Datos Maestros** [16]. Los Datos Maestros son entidades, relaciones y atributos críticos para las organizaciones, y a su vez clave para los BP y sistemas de aplicación [14].

La gestión de datos maestros (MDM) aporta soluciones a problemas tales como que los datos intercambiados entre las organizaciones sean los mismos pero reciban distinto nombre. También puede suceder que el mismo dato se utilice para identificar conceptos diferentes, o por el contrario que conceptos distintos reciben el mismo nombre [14].

Loshin en [14] define MDM (Master Data Management) como una recopilación de las mejores prácticas de gestión de datos orientadas a la integración de los datos. Estas prácticas se integran en las aplicaciones de negocio, gestión de

información, y herramientas para la gestión de datos que implementan procedimientos, servicios e infraestructura para el soporte de la captura, integración y compartición de los Datos Maestros. La MDM tiene asociados beneficios como el incremento de la calidad de los datos [15]. Una organización puede emprender acciones de evaluación y mejora de calidad de datos sin tener en cuenta mecanismos para la MDM. Sin embargo, la situación contraria no es posible, debido a que la MDM implica unos datos de calidad [13].

2.3 Familia ISO/IEC 8000-100

La premisa básica de la familia de estándares ISO/IEC 8000:2011, es usar un formato preestablecido para los Mensajes de Datos Maestros intercambiados en la comunicación entre aplicaciones. Además, se especifican los requisitos que deben cumplir dichos mensajes para incorporar información como parte de ese formato preestablecido sobre el nivel de calidad de los datos contenidos en el Mensaje de Datos Maestros. Un **Mensaje de Datos Maestros** es un mensaje usado para el intercambio de Datos Maestros entre organizaciones.

El objetivo específico de cada una de las partes de la familia de estándares ISO/IEC 8000:2011 se detallan a continuación:

- ISO 8000:100: describe los aspectos específicos de Datos Maestros para gestionarlos en sistemas de Gestión de Calidad de Datos [1].
- ISO 8000:102: describe el vocabulario relacionado con la calidad de datos maestros utilizado en las distintas partes del estándar [5].
- ISO 8000:110: establece las normas que se tienen que utilizar para la codificación de los mensajes de datos maestros (aspectos sintácticos, codificación semántica y requisitos) [6].
- ISO 8000:120: establece la forma en la que se puede añadir información sobre el ciclo de vida de los datos y la evolución de éstos (data provenance) [2].
- ISO 8000:130: establece cómo añadir información sobre el grado de **precisión** que tienen los datos [3].
- ISO 8000:140: establece cómo añadir información sobre el grado de **compleción** que tienen los datos [4].

3 I8K: una Implementación de ISO/IEC 8000-100

En esta sección se describe la Arquitectura de Servicios I8K, que implementa la familia de estándares ISO/IEC 8000-100. También se detallan los modelos de datos, una adaptación para el formato de los mensajes y el protocolo de comunicación para llevar a cabo el intercambio de Mensajes de Datos Maestros entre las aplicaciones e I8K y entre las diferentes aplicaciones que se intercambian datos.

3.1 Arquitectura de Servicios I8K

La **Arquitectura de Servicios I8K** está compuesta por los agentes necesarios para satisfacer los requisitos de las distintas partes del estándar ISO/IEC 8000:2011-100. Por tanto, I8K permite:

1. Formatear los mensajes de acuerdo al modelo de datos y
2. Añadir información sobre el nivel de calidad de los datos maestros contenidos en los mensajes, tomando como base el formato de los mensajes.

Dichos mensajes serán intercambiados entre las aplicaciones que necesiten añadir información sobre el nivel de calidad de datos. En la Fig. 2 se muestra un modelo de despliegue de la arquitectura. I8K está basada en una arquitectura MOM (Message-Oriented Middleware) [12].

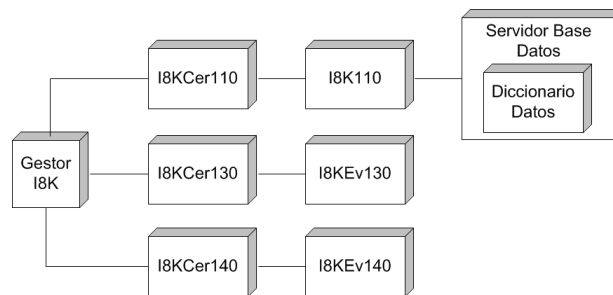


Fig. 2. Modelo de despliegue de la Arquitectura de Servicios I8K.

A continuación se detallan las funcionalidades específicas de cada uno de los diferentes agentes:

- *GestorI8K*: este agente es el encargado de recibir y procesar en primera instancia las peticiones que se realizan a la Arquitectura de Servicios I8K desde las aplicaciones que necesitan certificar sus datos. Hace de fachada a la arquitectura y delega las diferentes responsabilidades en los distintos agentes.
- *I8K.Cer110*: se encarga de la gestión y mantenimiento del Diccionario de Datos que requiere la parte 110.
- *I8K110*: se encarga de codificar y decodificar los distintos mensajes de datos maestros con respecto al modelo de datos que hemos creado de acuerdo a los requisitos específicos de la parte 110 del estándar.
- *I8K.Cer130*: es el encargado de añadir la información de certificación del nivel de **Precisión** de los datos maestros contenidos en el mensaje de datos maestros según especifican los requisitos de la parte 130 del estándar. Esta información de precisión es obtenida a partir de la evaluación correspondiente realizada por algún otro agente externo, o por el I8K.Ev130.
- *I8K.Cer140*: es el encargado de añadir la información de certificación del nivel de **Compleción** de los datos maestros contenidos en el mensaje de

datos maestros según especifican los requisitos de la parte 140 del estándar. Esta información de compleción es obtenida a partir de la evaluación correspondiente realizada por algún otro agente externo, o por el I8K.Ev140.

- *I8K.Ev130*: es el encargado de evaluar la **Precisión** de los datos maestros contenidos en el mensaje de datos maestros.
- *I8K.Ev140*: es el encargado de evaluar la **Compleción** de los datos maestros contenidos en el mensaje de datos maestros.

Además, todos los agentes añaden información sobre el ciclo de vida de los datos (data provenance) según los requisitos mostrados en ISO/IEC 8000-120. De esta forma, los mensaje de datos maestros que se intercambien las aplicaciones cumplen así el estándar ISO/IEC 8000:2011.

Para poder explicar los tipos de mensajes de datos maestros existentes, es necesario diferenciar las partes implicadas. Por un lado, tenemos las aplicaciones que hacen uso de las funcionalidades ofrecidas por I8K. Estas aplicaciones se dividen en dos: (1) AP* que son las aplicaciones que piden datos a una serie de proveedores de datos y que esperan las respuestas de estos proveedores para combinarlas y, (2) los proveedores de datos. Ambas aplicaciones, AP* y proveedores, además de intercambiarse mensajes entre ellos, se intercambian mensajes con I8K.

Los tipos de mensajes de datos maestros que se intercambian entre las aplicaciones y la Arquitectura de Servicios I8K viene detallados en la tabla 1.

Tipo	Descripción
I8K.COD-GE	Una aplicación necesita codificar un Mensaje de Datos Maestros.
I8K.CODED	I8K ha codificado un Mensaje de Datos Maestros.
I8K.DEC	Una aplicación necesita decodificar un Mensaje de Datos Maestros para poder entender su contenido.
I8K.DECODED	I8K ha decodificado un Mensaje de Datos Maestros.
I8K.COD-CR130	Una aplicación necesita codificar el mensaje y evaluar y certificar los datos maestros que contiene dicho mensaje de acuerdo al nivel de calidad de precisión.
I8K.COD-CR140	Una aplicación necesita codificar el mensaje y evaluar y certificar los datos maestros que contiene dicho mensaje de acuerdo al nivel de calidad de compleción.
I8K.COD-CR	Una aplicación necesita codificar el Mensaje de Datos Maestros, y evaluar y certificar los datos maestros que contiene dicho mensaje de acuerdo a los niveles de precisión y compleción.

Table 1. Tipos de Mensajes de Datos Maestros intercambiados entre las aplicaciones e I8K.

Por otro lado, los tipos de mensajes de datos maestros que se intercambian AP* y los proveedores vienen detallados en la tabla 2.

Tipo	Descripción
I8K.REQ	Una aplicación envía un mensaje de petición de datos a un proveedor de datos.
I8K.RES	Un proveedor de datos envía un mensaje de respuesta con los datos maestros solicitados por otra aplicación.

Table 2. Tipos de Mensajes de Datos Maestros enviados entre una aplicación y un proveedor de datos.

3.2 Modelo de Datos y formato de los mensajes de datos maestros

Los mensajes de datos maestros tienen que cumplir una **Sintaxis Formal**, que debe expresarse como un par (*propiedad, valor*). Según [6], el par representa una instancia de un valor específico junto con su valor codificado. Por ejemplo si el dato maestro *FechaInicio* está contenido dentro de un mensaje de datos maestros, para cumplir el estándar entonces en el mensaje aparecerá su valor codificado, por ejemplo, *HLJK*. Para ello, en la sección normativa de la parte 110, se requiere la creación de un **Diccionario de Datos (DD)**. La "propiedad" se refiere a los términos del vocabulario del dominio bajo estudio, y tanto los términos como los correspondientes valores tienen que estar almacenados en el DD.

Como parte de la implementación de I8K, se ha desarrollado un modelo de datos (que permite describir la sintaxis) para que dé soporte a nuestro propio diccionario de datos. El modelo de datos consta de los siguientes elementos:

- **Término:** este campo permite especificar los descriptores que se incluyen en el vocabulario. Por ejemplo *FechaInicio*
- **Estado:** representa si un término está activo o no dentro del DD.
- **Idioma:** especifica el idioma en el que el término está en el DD. El mismo término puede estar almacenado en diferentes idiomas.
- **Organización:** es la información de la organización que ha introducido en término en el DD.
- **Definición:** definición del término. Si el término está almacenado en varios idiomas, DD tendrá las correspondientes definiciones en los idiomas.
- **Identificador:** este campo contiene el valor del término codificado. Por tanto cuando un mensaje de datos maestros sea codificado el valor del campo término se sustituirá por el valor de este campo. Por ejemplo si el dato maestro *Destino* está contenido dentro de un mensaje de datos maestros, en el mensaje aparecerá su valor codificado, por ejemplo *DTNO*.
- **Nombre de la organización propietaria del término:** nombre de la organización que almacena el término.

Los Mensajes de Datos Maestros tienen que cumplir una **Sintaxis Formal** específica para poder almacenar la información de la calidad de los datos. Por lo que, tanto la AP* como los diferentes Proveedores de datos van a necesitar comunicarse con el GestorI8K con el fin de poder comunicarse entre ellos además de obtener las certificaciones de los niveles de calidad necesarios.

3.3 Protocolo de Comunicación

Para describir el funcionamiento de la Arquitectura I8K, se ha definido un **Protocolo de Comunicación**. Dicho protocolo regula el intercambio de mensajes de datos maestros entre las aplicaciones que usan la arquitectura, así como entre las propias aplicaciones.

En primer lugar, para que la AP* pueda solicitar datos a un Proveedor necesita seguir los pasos mostrados en la Fig. 3. El proceso comienza con el

paso 1, cuando AP* crea su petición. Para ello, crea un objeto que representa un mensaje de datos maestros. Este objeto encapsula como parte del mensaje: el tipo de mensaje, el vocabulario específico para el dominio, y los requisitos de gestión de calidad de datos. Esta información es necesaria para que el proveedor sea capaz de proporcionar los datos solicitados con el nivel adecuado de calidad de datos como parte de su respuesta. Una vez creado el mensaje, se lo envía al GestorI8K (paso 2) para que se lo codifique (mensaje tipo I8K.COD-GE según la tabla 1). A continuación, el gestor, en el paso 2.1, se encarga de codificar los datos y construir el nuevo mensaje codificado. Este mensaje de tipo I8K.CODED, el gestor I8K se lo envía de vuelta a AP* (paso 2.2) para que ésta se lo envíe al proveedor (paso 3) como un mensaje I8K-REQ (ver tabla 2).

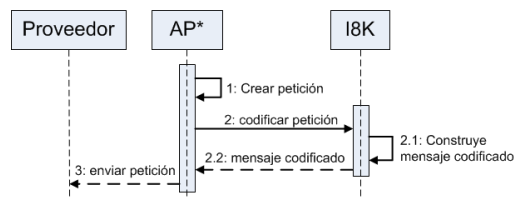


Fig. 3. Diagrama de secuencia a realizar por la aplicación.

En segundo lugar, para que los proveedores de datos puedan certificar sus datos, deben realizar la siguiente secuencia de pasos:

1. Recepción de petición de datos: el proveedor recibe una petición de datos por parte del AP* de tipo I8K.REQ.
2. Petición de decodificación al I8K y recepción : el proveedor reenvía el mensaje recibido al GestorI8K para que se lo decodifique y poder saber qué datos le esta pidiendo el AP*.
3. Recepción de mensaje decodificado y realización de consultas: el proveedor recibe los datos decodificados y pasa a realizar las consultas y operaciones oportunas para generar el resultado a devolver a la AP*.
4. Petición de codificación y certificación al I8K: el proveedor envía el resultado generado al gestor I8K.
5. Codificación y certificación: el gestor I8K codifica los datos recibidos por parte del proveedor. Además, si se ha solicitado certificación por parte del AP*, entonces procede también a certificar los datos según las partes 130 y/o 140.
6. Recepción de mensaje codificado y/o certificado y envío a AP*: el proveedor una vez que recibe el mensaje con los datos codificados y certificados (si fuera necesario), se lo reenvía a AP* como respuesta (mensaje tipo I8K.RES)

Por otra parte, la Arquitectura I8K funciona tal como se describe a continuación. Cada vez que la Arquitectura I8K recibe un mensaje de datos maestros, es a través del agente GestorI8K. El GestorI8K se encarga en primera instancia de procesarlo en función de la información almacenada en su cabecera. En este

procesamiento del mensaje, se extrae el tipo de mensaje que es y se reenvía al agente encargado de la realización de la tarea que indica dicho tipo.

Según el tipo de mensaje, las acciones que se pueden hacer y los agentes encargados de hacerlas son los siguientes:

- *Codificar/Decodificar (Agente I8KCer110)*: este agente procesa la cabecera del mensaje y se lo reenvía al agente I8K110 que procesa la cabecera para ver la sintaxis que tiene que utilizar para codificar y decodificar el mensaje de datos maestros.
- *Certificar el nivel de precisión (Agente I8KCer130)*: envía el mensaje de datos maestros al agente I8KEv130. Este último evalúa los datos contenidos en el mensaje, y le envía el resultado de la evaluación al I8KCer130 que será quien se encargue de certificar que los datos contenidos en ese mensaje tenían el citado nivel de precisión.
- *Certificar el nivel de compleción (Agente I8KCer140)*: este agente envía el mensaje de datos maestros al agente I8KEv140. Este último evalúa los datos contenidos en el mensaje, y le envía el resultado de la evaluación al I8KCer140 que será quien se encargue de dar fe (certificar) que los datos contenidos en ese mensaje tenían el citado nivel de compleción.

Cuando los agentes I8KCer110, I8KCer130 e I8KCer140 terminan el procesamiento requerido del mensaje de datos maestros, devuelven el mensaje de datos maestros incluyendo la correspondiente información del nivel de calidad requerido al GestorI8K. El GestorI8K lo devuelve a su vez a la aplicación que ha solicitado cualquiera de las funcionalidades, que será quien se encargue de procesar la nueva información y tomar decisiones al respecto.

4 AP* Mejorada para contemplar Calidad de Datos

Para que el proceso de negocio descrito en el apartado 1.1 tenga en cuenta los aspectos de calidad de datos descritos hasta ahora, es necesario realizar ciertas mejoras. Esta mejora se compone de dos partes fundamentales: los cambios en las actividades basadas en servicios a combinar que proporcionan los datos requeridos por la aplicación, a la que llamamos AP*, y los cambios en la propia aplicación AP*.

4.1 Adaptación de la Actividades para el soporte de Calidad de Datos

Como premisa inicial, es necesario que las organizaciones encargadas de las diferentes actividades a combinar, se comprometan y garanticen que van a certificar sus datos a través de la arquitectura I8K. Estas actividades generalmente basadas en servicios, son los proveedores de datos de la aplicación, por lo que a partir de la petición que reciben de esta, generan una respuesta que contiene el resultado que ofrece a través de sus servicios. La primera modificación viene en la adaptación del tipo de mensaje que recibe. Mientras que antes eran los datos

concretos necesarios para generar el resultado, ahora es un mensaje codificado según la parte 110, que a parte de los datos concretos, incluyen información sobre los niveles de calidad requeridos. Por tanto, las organizaciones tienen que adaptar sus servicios para realizar los pasos descritos en el apartado 3.3. En el ejemplo ilustrativo (Fig. 1), las actividades de Búsqueda de Vuelo, Búsqueda de Hotel y Búsqueda de Alquiler de Coche se han adaptado para realizar la secuencia de pasos descritos en el apartado 3.3, y poder certificar así los datos que envían a la aplicación.

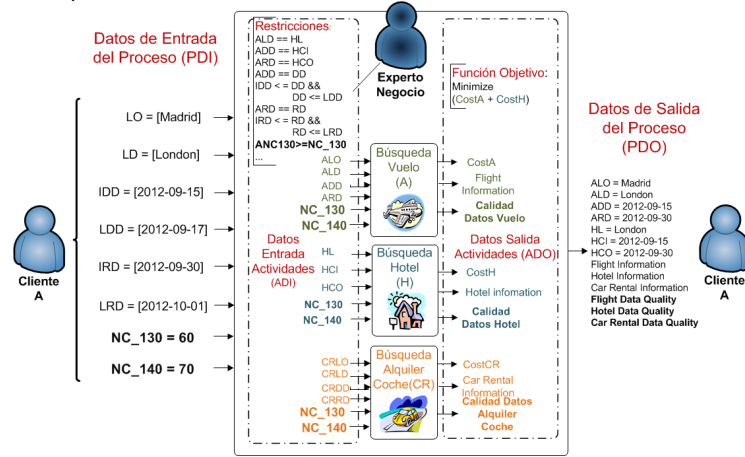


Fig. 4. Datos y Restricciones del Organizador de Viaje.

4.2 Cambios en la AP*

La AP* presentada en la Sección 1.1 se encarga de coordinar y enviar los datos a las diferentes actividades (también llamadas proveedores de datos) y luego comprobar y combinar los valores devueltos por las mismas para obtener el objetivo global. Además, esta AP* es la encargada de comprobar y garantizar que todas las restricciones que relacionan los datos se cumplan. Por tanto, es necesario realizar ciertas modificaciones para poder contemplar, cubrir y garantizar a los usuarios, los niveles de calidad adecuados.

En primer lugar, hay que solicitar al usuario los niveles de calidad que desea y que va a exigir a los servicios, por lo que se introducen dos nuevos datos de entrada al proceso: nivel deseado de calidad de la parte 130 (al que llamaremos, NC_130) y nivel deseado de calidad de la parte 140 (al que llamaremos NC_140).

Además, la aplicación tiene que incorporar nuevas restricciones que incluyan la calidad de datos. En el ejemplo utilizado, se introducen en concreto tres restricciones: (1) los niveles de calidad 130 y 140 devueltos por la actividad que busca el vuelo tienen que ser mayores o iguales a los exigidos por el usuario, (2) lo mismo para la búsqueda de hotel y (3) lo mismo para la búsqueda del alquiler de coche, en resumen:

$$\forall \text{Actividad} \in \text{Combinacion} \{ \{ NC_{130} \leq \text{Actividad}.NC_{130} \wedge NC_{140} \leq \text{Actividad}.NC_{140} \}$$

Estas restricciones ayudan a la toma de decisiones de la AP*, ya que condicionan el resultado global devuelto al usuario. Si los datos devueltos por una actividad no cumplen los niveles de calidad exigidos por el usuario, entonces no se tendrán en cuenta y se descartarán como parte de una posible solución global.

Por último, AP* debe seguir los pasos mostrados en la Fig. 3 para poder enviar peticiones a los proveedores y obtener los niveles de calidad de los datos devueltos por los mismos.

Los cambios realizados en el ejemplo ilustrativo se muestran en la Fig. 4. Se han incluido los parámetros de entrada y las restricciones necesarias en el proceso para que tengan en cuenta la calidad de datos. Hasta ahora, si los resultados eran válidos, entonces AP* combinaba dichos resultados para generar una solución, o lo que es lo mismo, el paquete de viajes. Por ejemplo, la aplicación comprobaba que el precio devuelto por el proveedor de vuelos era un valor positivo para combinarla con los resultados del hotel y del alquiler de coche. Ahora, no sólo comprueba la validez de los resultados, sino que comprueba los niveles de calidad de los mismos, de tal forma, que si no tienen el nivel de calidad adecuado, entonces no se combina.

5 Conclusiones

Una forma de solventar las amenazas que pueden producirse debido a datos incorrectos en la combinación de servicios, es estableciendo los niveles de calidad de los datos que se intercambian. La familia de estándares ISO/IEC 8000-100 proponen usar un formato preestablecido para el intercambio de datos y los requisitos que deben cumplir dichos mensajes incorporando la información sobre el nivel de calidad de los datos.

En este trabajo se propone una arquitectura de servicios, llamada I8K, que ofrece la funcionalidad propuesta por dicha familia de estándares. Esta arquitectura proporciona el soporte necesario para evaluar y certificar los datos de acuerdo a los niveles de calidad de precisión y completión. Además, proporciona el diccionario de datos necesario para el intercambio de mensajes. Junto a esto, se ha diseñado un protocolo de comunicación para el intercambio de mensajes entre las aplicaciones y la arquitectura I8K. Por último, esta propuesta se ha aplicado a un ejemplo donde se combinan diferentes actividades con el fin de incluir los aspectos de calidad de los datos. Esta aplicación se encarga de combinar diferentes actividades para obtener un objetivo global. En esta aplicación los datos toman una especial relevancia, ya que condicionan el resultado ofrecido a los usuarios. Con nuestra arquitectura I8K, se puede garantizar que los datos ofrecidos a los usuarios cumplen los niveles adecuados de precisión y completión.

Agradecimientos. Este trabajo ha sido parcialmente financiado por los proyectos de investigación MAGO (TIN2009-13718-C02-01), TDiaCO-BPMS (TIN2009-13714), subvencionados por el Ministerio de Ciencia y Tecnología de España y por el Fondo Europeo de Desarrollo Regional (ERDF/FEDER).

References

1. ISO/DIS 8000-100: Master Data: Exchange of characteristic data: Overview. ISO/IEC, 2011.
2. ISO/DIS 8000-120: Master Data: Exchange of characteristic data: Provenance. ISO/IEC, 2011.
3. ISO/DIS 8000-130: Master Data: Exchange of characteristic data: Accuracy. ISO/IEC, 2011.
4. ISO/DIS 8000-140: Master Data: Exchange of characteristic data: Completeness. ISO/IEC, 2011.
5. ISO/IEC 8000-102: Master Data: Exchange of characteristic data: Vocabulary. ISO/IEC, 2011.
6. ISO/IEC 8000-110: Master Data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. ISO/IEC, 2011.
7. Carlo Batini and Monica Scannapieca. *Data Quality: Concepts, Methodologies and Techniques*. Springer-Verlag Berlin Heidelberg, 2006.
8. Lucas Bordeaux, Youssef Hamadi, and Horst Samulowitz. Experiments with massively parallel constraint solving. In *IJCAI*, pages 443–448, 2009.
9. Ismael Caballero, Angelica Caro, Coral Calero, and Mario Piattini. Iqm3: Information quality management maturity model. *J. UCS*, 14(22):3658–3685, 2008.
10. ECCMA. Electronic Commerce Code Managements Association.
11. Redouane Ezzahir, Christian Bessiere, Imade Benelallam, and Mustapha Belaisaoui. Dynamic backtracking for distributed constraint optimization. pages 901–902, 2008.
12. Ian Gorton. *Essential Software Architecture*. Springer, 2011.
13. Andy Hayler. Data quality essential to master data management. 2012.
14. David Loshin. *Master Data Management*. Elsevier, 2009.
15. Boris Otto and Andreas Reichert. Organizing master data management: findings from an expert survey. proceedings of the 2010 acm symposium on applied computing. 2010.
16. Boris Otto and Alexander Schmidt. Enterprise Master Data Architecture: Design Decision and Options. 2010.
17. Mike P. Papazoglou and Willem-Jan van den Heuvel. Service oriented architectures: approaches, technologies and research issues. *VLDB J.*, 16(3):389–415, 2007.
18. Luisa Parody, María Teresa Gómez-López, and Rafael M. Gasca. Extending bpmn 2.0 for modelling the combination of activities that involve data constraints. pages 68–82, 2012.
19. Luisa Parody, María Teresa Gómez-López, Rafael Martínez Gasca, and Angel Jesus Varela-Vaca. Improvement of Optimization Agreements in Business Processes involving Web Services. *Communications of the IBIMA*, 2012, 2012.
20. Thomas C. Redman. *Data Quality for the Information Age*. Artech House Inc, 1996.
21. Thomas C. Redman. *Data Driven: Profiting from Your Most Important Business Asset*. Harvard Business Press, 2008.
22. Wil M. P. van der Aalst, Arthur H. M. ter Hofstede, and Mathias Weske. Business process management: A survey. In *Business Process Management*, pages 1–12, 2003.
23. Mathias Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, 2007.
24. Makoto Yokoo and Katsutoshi Hirayama. Algorithms for distributed constraint satisfaction: A review. In *AAAI*, volume 3, pages 198–212, 2000.